UNCLASSIFIED

AD 410291

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION, ALEXANDRIA, VIRGINIA

UNCLASSIFIED

DEPARTMENT OF STATISTICS

University of Wisconsin
Madison, Wisconsin

TECHNICAL REPORT NO. 15

February, 1963

A BAYESIAN APPROACH TO SOME BEST POPULATION PROBLEMS

Irwin Guttman and George C. Tiao

This research was supported in part by the United States Navy through Office of Naval Research, under Contract Nonr-1202(17), Project NR 042 222. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Also issued as Systems Formulation and Methodology Workshop Paper No. 6304 by the Social Systems Research Institute, University of Wisconsin.

A BAYESIAN APPROACH TO SOME BEST POPULATION PROBLEMS

Irwin Guttman and George C. Tiao

I. Introduction and Summary

There have been several papers recently in the literature devoted to the subject of selecting a "best" population -- see for example Gupta and Sobel (1962), Guttman (1961) and others. In these works, the problem was analyzed from the sampling theory point of view. Except for some simple cases, this approach frequently leads to the problem of eliminating nuisance parameters. And, unless rather strong assumptions about certain of the parameters involved are made, the problem usually becomes intractable.

In this paper, we consider certain best population problems adopting a Bayesian approach. One main advantage of such an approach is that we are able to arrive at satisfactory decision procedures in the presence of nuisance parameters. Indeed, the use of Bayes' theorem allows one to analyse best population problems from many points of view. The plan of this paper is as follows:

In section 2, a general description of best population problems is first outlined. The criterion for "bestness" is regarded as a "utility" function of the statistician. The decision procedure then adopted is based upon the principle of maximizing posterior expected utility. In sections 3 and 4, we discuss the application of this procedure when sampling from normal populations and exponential populations, respectively. In both cases, the criterion defining the best population is taken to be the coverage of the population considered in a certain given interval. The procedures are shown in section 5 to be consistent. Further, in section 6, we extend the decision analysis by considering the posterior distribution of the criterion. This extension enables us to propose other decision procedures which may be more appropriate in certain situations than that resulting from the principle of maximizing posterior expected utility. Some examples are illustrated in detail. Finally, in section 7, we discuss briefly some different types of best population problems.

II. Statement of the Problem

We are given a collection Π of k populations $\Pi_1, \Pi_2, \ldots, \Pi_i, \ldots, \Pi_k$ in which there exists a so-called "best" population, where the population is "best" according to a specific definition of the following kind.

Suppose the Π_i's are distributed with probability density function $f(x|\theta_i)$, where θ_i is possibly vector-valued. Consider a real valued function of θ_i, $h_i = g(\theta_i)$, where g is known. Then the "best" population is defined to be that population which has largest value among the $h_i = g(\theta_i)$, $i = 1, \ldots, k$. We assume there is an ordering of the $h_1, \ldots, h_k$ into

$$h_{[1]} < h_{[2]} \ldots < h_{[k-1]} < h_{[k]} .$$

As we are interested in determining which of the k populations has as its value of h, the value $h_{[k]}$, and in accordance with the Bayesian approach, we may therefore regard the function $h_i = g(\theta_i)$ as the utility function of the statistician re the population Π_i. The statistician's procedure will be as follows:

From each of the Π_i select a random sample $\underline{y}_i$ of size n_i. Determine the a posteriori distribution $p(\theta_i|\underline{y}_i)$ of θ_i given the sample and hence find $E(h_i|\underline{y}_i)$, the expectation of h_i, $i = 1, \ldots, k$. The statistician will then choose as the "best" population that population which assumes the maximum value of these expectations. That is, the one with value

$$\max_i E(h_i|\underline{y}_i) = \max_i \int h_i \, p(\theta_i|\underline{y}_i) \, d\theta_i \ .$$

In sections 3 and 4, we apply this approach to two particular cases. We define $h_i = g(\theta_i) = \int_{a_1}^{a_2} f(x|\theta_i) \, dx$, where a_1 and a_2 are known, and consider the cases:

(a) f is the normal density, and

(b) f is the exponential density.

This problem with h_i defined as above, has been considered by Guttman (1961) from the non-Bayesian point of view. We note that several of the more interesting cases which lead to considerable mathematical difficulties when the non-Bayesian approach is used are readily solvable adopting the above-mentioned procedure (see below).

III. The Case of the Normal Density

When the populations Π_i are normally distributed, we have that

$$h_i = g(\mu_i, \sigma_i) = \int_{a_1}^{a_2} \frac{1}{\sqrt{2\pi}\,\sigma_i} \quad \exp\left\{- \frac{1}{2\sigma_i^2} (y-\mu_i)^2 \right\} dy.$$

Let a sample $\underline{y}_i = (y_{i1}, \ldots, y_{in_i})$ be taken from Π_i. We have for the likelihood function:

$$\ell(\mu_i, \sigma_i|\underline{y}_i) \propto \sigma_i^{-n_i} \quad \exp\left\{- \frac{1}{2\sigma_i^2} \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2 \right\}$$

(3.1)

$$\propto \sigma_i^{-n_i} \quad \exp\left\{- \frac{1}{2\sigma_i^2} \, [n_i \, s_i^2 + n_i \, (\bar{y}_i - \mu_i)^2] \right\}$$

where $\bar{y}_i$ and s_i^2 are the maximum likelihood estimators of μ_i and σ_i^2, respectively.

Assume that we are in a situation where little is known a priori re the values of (μ_i, σ_i), for all i. In other words, we are saying that the information we have re (μ_i, σ_i) comes primarily from the sample $\underline{y}_i$. We may then adopt the approach used by Jeffreys (1961), Savage (1961), and Box and Tiao (1962), and assume that μ_i and $\log \sigma_i$ are independent and locally uniformly distributed a priori. That is,

$$(3.2)\qquad \begin{aligned} p(\mu_i) &\propto k_1 \\ p(\log \sigma_i) &\propto k_2 \text{ or } p(\sigma_i) \propto \frac{1}{\sigma_i} \end{aligned}$$

all i.

Using (3.2), the joint posterior distribution of (μ_i, σ_i) is:

$$(3.3)\qquad p(\mu_i, \sigma_i|\underline{y}_i) = p_1(\sigma_i|s_i)\ p_2(\mu_i|\sigma_i, \bar{y}_i)$$

where

$$(3.4a)\qquad p_1(\sigma_i|s_i) = 2\left\{\Gamma\left(\frac{n_i-1}{2}\right)\right\}^{-1}\left\{\left(\frac{n_i s_i^2}{2}\right)^{\frac{n_i-1}{2}}\right\}\sigma_i^{-n_i}\exp\left\{-\frac{n_i s_i^2}{2\sigma_i^2}\right\}$$

and

$$(3.4b)\qquad p_2(\mu_i|\sigma_i, \bar{y}_i) = \left\{\left(\frac{n_i}{2\pi\sigma_i^2}\right)\right\}^{\frac{1}{2}}\exp\left\{\left(-\frac{n_i}{2\sigma_i^2}\right)(\bar{y}-\mu_i)^2\right\}.$$

We note from (3.3) that the adoption of the prior distributions (3.2) amounts to saying that the joint posterior distribution of μ_i and $\log \sigma_i$ is approximated by the likelihood function (3.1).

Consider first the case where $a_1 = -\infty$. Then, the a posteriori expectation of h_i is

$$\begin{aligned} E(h_i|\underline{y}_i) &= \int_{-\infty}^{\infty}\int_0^{\infty} h_i\ p_1(\sigma_i|s_i)\ p_2(\mu_i|\bar{y}_i, \sigma_i)\ d\sigma_i\ d\mu_i \\ &= k_3\int_{-\infty}^{\infty}\int_0^{\infty}\int_{-\infty}^{a_2}\sigma_i^{-(n_i+2)}\exp\left\{-\frac{1}{2\sigma_i^2}\left[(\mu_i-z)^2+n_i s_i^2+n_i(\bar{y}_i-\mu_i)^2\right]\right\} \\ &\qquad dz\ d\sigma_i\ d\mu_i. \end{aligned}$$

It is clear that we may reverse the order of integration. On integrating out σ_i and μ_i we obtain

$$(3.5) \quad E(h_i|\underline{y}_i) = \frac{\Gamma(\frac{n_i}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n_i-1}{2})}[(n_i+1)s_i^2]^{-\frac{1}{2}}\int_{-\infty}^{a_2}\left\{1+\frac{(\bar{y}_i-z)^2}{(n_i+1)s_i^2}\right\}^{-\frac{n_i}{2}}dz.$$

That is, we have the following remarkably simple result

$$(3.6) \quad E(h_i|\underline{y}_i) = F_{n_i-1}(t_i)$$

where F is the cumulative distribution function of the Student-t variable with (n_i-1) degrees of freedom and $t_i = (\frac{n_i-1}{n_i+1})^{\frac{1}{2}}(\frac{a_2-\bar{y}_i}{s_i})$.

Similarly, if a_1 is finite, we find that

$$(3.7) \quad E(h_i|\underline{y}_i) = F_{n_i-1}[t_i^{(2)}] - F_{n_i-1}[t_i^{(1)}],$$

where F is, as before, the Student-t cumulative distribution with (n_i-1) degrees of freedom and

$$(3.8) \quad t_i^{(1)} = (\frac{n_i-1}{n_i+1})^{\frac{1}{2}}(\frac{a_1-\bar{y}_i}{s_i}), \quad t_i^{(2)} = (\frac{n_i-1}{n_i+1})^{\frac{1}{2}}(\frac{a_2-\bar{y}_i}{s_i}).$$

The statistician's procedure is now clear. If one behaves as a maximizer of his expected utility -- see for example Luce and Raiffa (1957), and Raiffa and Schlaifer (1961), he computes the expectations (3.6) or (3.7) depending on whether $a_1 = -\infty$ or a_1 is finite, respectively, for each population. He then chooses as the best population the one with the sample giving the maximum value of these expectations.

In the case where $a_1 = -\infty$ (this will be called the one-sided case), we note that if $n_i = n$, all i, then this selection procedure is equivalent to choosing the population which has as value of $\frac{a_2-\bar{y}_i}{s_i}$, the value

$\max\limits_{j} \frac{a_2-\bar{y}_j}{s_j}$. This is an intuitively pleasing result for the following reason. We are interested in the best population, that is, the population which has largest value of

$$h_i = g(\mu_i, \sigma_i) = \int_{-\infty}^{a_2} \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma_i^2}(x-\mu_i)^2\right\} dx$$

$$= \int_{-\infty}^{\frac{a_2-\mu_i}{\sigma_i}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{t^2}{2}\right\} dt,$$

and since this is a monotone increasing function of the upper limit, say $\tau_i = \frac{a_2-\mu_i}{\sigma_i}$, the best population is that with largest value of τ_i. It is intuitively evident that an estimate of τ_i should be based on $\frac{a_2-\bar{y}_i}{s_i}$ and hence, that the largest value of $\frac{a_2-\bar{y}_i}{s_i}$ should be indicative of the best population.

IV. The Case of the Exponential Density

In this section we discuss the situation in which the populations Π_i have exponential distributions. That is, their probability density functions $f(y|\theta_i)$ take the form:

$$(4.1) \quad f(y|\mu_i, \sigma_i) = \begin{cases} \frac{1}{\sigma_i} \exp\left\{-\left(\frac{y-\mu_i}{\sigma_i}\right)\right\}, & y \geq \mu_i \\ 0 & \text{otherwise.} \end{cases}$$

We take up the one-sided case first, namely, we let $a_2 = a$, and $a_1 = -\infty$. Then, h_i becomes

$$(4.2) \quad h_i = g(\mu_i, \sigma_i) = \int_{-\infty}^{a} f(y|\mu_i, \sigma_i)\, dy.$$

Let a sample $\underline{y}_i$ of size n_i be taken from Π_i, $i = 1, \ldots, k$, and the observations in $\underline{y}_i$ be ordered, that is we let $\underline{y}_i = (y_{i1}, \ldots, y_{in_i})$ where here $y_{i1} < y_{i2} < \ldots < y_{in_i}$. The likelihood function is given by:

$$(4.3)\qquad \ell(\mu_i, \sigma_i|\underline{y}_i) = \sigma_i^{-n_i} \exp\left\{-\frac{1}{\sigma_i} \sum_{j=1}^{n_i} (y_{ij} - \mu_i)\right\}$$

$$= \sigma_i^{-n_i} \exp\left\{-\frac{1}{\sigma_i}[(n_i-1)\, w_i + n_i(y_{i1}-\mu_i)]\right\}$$

where $w_i = (n_i-1)^{-1} \sum_{j=2}^{n_i} (y_{ij}-y_{i1})$.

It is to be noted that for all $\sigma_i > 0$ the likelihood function (4.3) is a monotonic increasing function of μ_i in the interval $(-\infty, y_{i1})$ and vanishes outside this interval. Two interesting cases may now arise, namely, (1) $y_{i1} < a$, and (2) $y_{i1} > a$.

For a given population Π_i, suppose case (1) occurs. This implies that $\mu_i < a$ and hence (4.2) becomes

$$(4.4)\qquad h_i = 1 - \exp\left\{-\frac{a-\mu_i}{\sigma_i}\right\} .$$

As in section 3, we assume that the prior distributions for μ_i and $\log \sigma_i$ are locally uniform. The joint posterior distribution of μ_i and σ_i then takes the form:

$$(4.5)\qquad p(\mu_i, \sigma_i|\underline{y}_i) = p_3(\sigma_i|w_i)\, p_4(\mu_i|\sigma_i, y_{i1})$$

where

$$(4.6a)\qquad p_3(\sigma_i|w_i) = \frac{\left\{(n_i-1)w_i\right\}^{(n_i-1)}}{\Gamma(n_i-1)} \sigma_i^{-n_i} \exp\left\{-\frac{(n_i-1)w_i}{\sigma_i}\right\},\ \sigma_i > 0$$

and

$$(4.6b)\qquad p_4(\mu_i|\sigma_i, y_{i1}) = \frac{n_i}{\sigma_i} \exp\left\{-\frac{n_i}{\sigma_i}(y_{i1}-\mu_i)\right\} , \qquad \mu_i < y_{i1}.$$

Hence, the expected utility is:

$$(4.7)\qquad \begin{aligned} E(h_i|\underline{y}_i) &= \int_{-\infty}^{y_{i1}} \int_0^{\infty} h_i\, p_3(\sigma_i|w_i)\, p_4(\mu_i|\sigma_i, \underline{y}_{i1})\, d\sigma_i\, d\mu_i \\ &= 1 - \frac{n_i}{n_i+1}\left[1 + \frac{a-y_{i1}}{(n_i-1)w_i}\right]^{-(n_i-1)}. \end{aligned}$$

We again note that this result (4.7) is intuitively pleasing. To begin with, we are interested in the population with largest value of h_i as given in (4.4), which is a monotone increasing function of $\frac{a-\mu_i}{\sigma_i}$. Hence, the population with largest value of $\frac{a-\mu_i}{\sigma_i}$ is the best population. It seems natural, therefore, that a selection procedure for the best population should be based on $\frac{a-y_{i1}}{w_i}$. In the special case $n_i = n$, this leads us to choose that population which yields the sample with largest value of $\frac{a-y_{i1}}{w_i}$. The above reasoning is borne out in expression (4.7), since it is a monotonic increasing function of $\frac{a-y_{i1}}{w_i}$.

Suppose now that for a population Π_i, case (2) arises; then there is no information from the sample to tell us which of the situations (i) $a < \mu_i$ or (ii) $\mu_i < a$ obtains. If, in fact, our prior information includes the knowledge that $a < \mu_i$, then the utility for that population is zero and this population should be dropped from further consideration.

On the other hand, suppose that a priori little is known about μ_i, except that it cannot exceed "a", and hence the utility h_i is again given by (4.4). As before, one can approximate the posterior distribution of μ_i and $\log \sigma_i$ by the likelihood function, but with the extra restriction $\mu_i < a$. We then have for the joint posterior distribution of μ_i and σ_i,

(4.8) $$p(\mu_i, \sigma_i|\underline{y}_i) = p_5(\mu_i|\sigma_i, a)\, p_6(\sigma_i|\bar{y}_i, a)$$

where

(4.9a) $$p_5(\mu_i|\sigma_i, a) = \frac{n_i}{\sigma_i} \exp\left\{-\frac{n_i}{\sigma_i}(a-\mu_i)\right\}, \qquad \mu_i < a,$$

and

(4.9b) $$p_6(\sigma_i|\bar{y}_i, a) = \frac{[n_i(\bar{y}_i-a)]^{(n_i-1)}}{\Gamma(n_i-1)} \sigma_i^{-n_i} \exp\left\{-\frac{n_i(\bar{y}_i-a)}{\sigma_i}\right\}$$

with $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$.

Using (4.8) we find that the expected utility is simply

$$E(h_i|\underline{y}_i) = \frac{1}{n_i+1} . \tag{4.10}$$

We note that the expected utility in (4.10) approaches zero as n_i tends to infinity. This is as it should be for here $\mu_i < a < y_{i1}$, and therefore, as $n_i \to \infty$, y_{i1} converges (in probability) to μ_i with the result that the utility (4.4) must approach zero. This has an interesting implication; when $a < y_{i1}$ and for a large sample, we have that the posterior expectation for our utility is either approximately (if $\mu_i < a$) or exactly (if $a < \mu_i$) zero. Thus, for a large sample, irrespective of whether or not we know that μ_i is greater or less than "a", the population Π_i should be dropped from further consideration.

Further, suppose the sample sizes are small and that the intermediate situation occurs, that is, $y_{i1} < a$ for some of the Π_i and $a < y_{i1}$ for the remaining ones. In the event that $n_i = n$, all i, the expression in (4.10) is always less than that in (4.7). The statistician's procedure is then clear -- he will immediately drop from further consideration those populations with $a < y_{i1}$ and select the best population from the remaining ones.

We now consider the "two-sided" case for the exponential, that is, we will be interested in the population having largest value among the quantities:

$$h_i = g(\mu_i, \sigma_i) = \int_{a_1}^{a_2} \frac{1}{\sigma_i} \exp\left\{- \frac{1}{\sigma_i} (x-\mu_i)\right\} dx \tag{4.11}$$

where a_1 and a_2 are known constants.

For a given population Π_i, we shall only be concerned with the following two cases:

(4.12a) $\mu_i < y_{i1} < a_1 < a_2$, and

(4.12b) $\mu_i < a_1 < y_{i1} < a_2$.*

Paralleling the above development, we find that the expected utilities are

(4.13) $$E(h_i|\underline{y}_i) = \frac{n_i}{n_i+1}\left[\left\{1+\frac{a_1-y_{i1}}{(n_i-1)w_i}\right\}^{-(n_i-1)} - \left\{1+\frac{a_2-y_{i1}}{(n_i-1)w_i}\right\}^{-(n_i-1)}\right]$$

under (4.12a), and

(4.14) $$E(h_i|\underline{y}_i) = \frac{n_i}{n_i+1}\left[1 - \left\{1+\frac{a_2 - a_1}{n_i(\bar{y}_i-a_1)}\right\}^{-(n_i-1)}\right]$$

when (4.12b) holds.

For the two-sided case, the procedure will be: Upon taking a sample of size n_i from each of the Π_i, compute the expectation in (4.13) or (4.14) according to whether condition (4.12a) or (4.12b) obtains. Then choose as the best population the one which has the largest expectation.

V. Consistency Properties of the Procedure

We show in this section that the procedures developed in sections 3 and 4 have a very interesting property, namely that of consistency.

Consider first sampling from normal populations (section 3). For the two-sided case, the a posteriori expectation of the utility h_i is given by (3.7). This is the area between $t_i^{(1)}$ and $t_i^{(2)}$ under a Student-t distribution with (n_i-1) degrees of freedom. As n_i tends to infinity, we see from (3.7) and (3.8) that this expectation approaches the value

---

*As in the one-sided case, we are assuming here that we have a priori information that μ_i is less than a_1.

$$(5.1) \qquad \lim_{n_i \to \infty} E(h_i|\underline{y}_i) = \int_{\frac{a_1-\bar{y}_i}{s_i}}^{\frac{a_2-\bar{y}_i}{s_i}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$$

where, of course, $\bar{y}_i$ and s_i are the __observed__ values of the mean and standard deviation of an infinitely large sample.

From the posterior distribution $p(\mu_i, \sigma_i|\underline{y}_i)$ in (3.3), it is straightforward to verify that the second moment of h_i is:

$$(5.2) \qquad E(h_i^2|\underline{y}_i) = \frac{1}{2\pi(1-\rho_i^2)^{\frac{1}{2}}} \int_{t_i^{(1)}}^{t_i^{(2)}} \int_{t_i^{(1)}}^{t_i^{(2)}} \left[1+ \frac{\gamma_1^2+\gamma_2^2-2\gamma_1\gamma_2\rho_i}{(n_i-1)(1-\rho_i^2)}\right]^{-\frac{n_i+1}{2}} d\gamma_1 d\gamma_2$$

with $\rho_i = \frac{1}{n_i+1}$.

The expression in (5.2) is a bivariate t-integral -- see Dunnett and Sobel (1954). As $n_i \to \infty$, ρ_i will tend to zero and the above integral approaches the following limit:

$$(5.3) \qquad \lim_{n_i \to \infty} E(h_i^2|\underline{y}_i) = \int_{\frac{a_1-\bar{y}_i}{s_i}}^{\frac{a_2-\bar{y}_i}{s_i}} \int_{\frac{a_1-\bar{y}_i}{s_i}}^{\frac{a_2-\bar{y}_i}{s_i}} \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(t^2+u^2)\right\} dt\, du$$

which is of course the square of (5.1). That is to say, in the limit the variance of h_i approaches zero. Thus, we have the important result that the utility h_i converges (in probability) to the value (5.1). Since the best population is the one with the largest value of the quantities in (5.1), $i = 1, \ldots, k$, our procedures described in section 3 insures that for large samples the best population will always be chosen. It will be seen that this large sample property also holds for the one-sided case, i.e. with $a_1 = \infty$.

Turning to sampling from the exponential (section 4), we now demonstrate the consistency property for the one-sided problem when y_{i1} is less than "a". Here, the utility h_i is given in (4.4). It is straightforward to verify that the r^{th} moment of h_i is

$$\text{(5.4)}\quad E(h_i^r|\underline{y}_i) = \sum_{\ell=0}^{r} \binom{r}{\ell} (-1)^{\ell} \left(\frac{n_i}{n_i+\ell}\right) \left\{1+\frac{\ell(a-y_{i1})}{(n_i-1)w_i}\right\}^{-(n_i-1)}$$

where we recall that y_{i1} is the first order statistic of the sample $\underline{y}_i$ and $w_i = \frac{1}{n_i-1}\sum_{j=2}^{n_i}(y_{ij} - y_{i1})$. In particular, by setting $r = 1$ and $r = 2$, we have that:

$$\text{(5.5a)}\quad E(h_i|\underline{y}_i) = 1 - \frac{n_i}{n_i+1}\left[1 + \frac{a-y_{i1}}{(n_i-1)w_i}\right]^{-(n_i-1)}$$

and

$$\text{(5.5b)}\quad \text{Var}\,(h_i|\underline{y}_i) = \frac{n_i}{n_i+2}\left[1 + \frac{2(a-y_{i1})}{(n_i-1)w_i}\right]^{-(n_i-1)} - \left(\frac{n_i}{n_i+1}\right)^2 \left[1+ \frac{a-y_{i1}}{(n_i-1)w_i}\right]^{-2(n_i-1)}.$$

The expression (5.5a) is of course the expected utility as given in (4.7). Now, as $n_i \to \infty$, the expressions (5.5a) and (5.5b) tend to, respectively,

$$\text{(5.6a)}\quad \lim_{n_i \to \infty} E(h_i|\underline{y}_i) = 1 - \exp\left\{-\frac{a-y_{i1}}{w_i}\right\}$$

and

$$\text{(5.6b)}\quad \lim_{n_i \to \infty} \text{Var}\,(h_i|\underline{y}_i) = 0.$$

In a similar way one can show for the other cases discussed in section 4 that the utility functions involved are such that they converge in probability to finite limits.

VI. Distribution Theory of the Utilities $h_i = c(\theta_i)$

Since the utility h_i is a function of θ_i, then once we have the posterior distribution of θ_i, we can, in principle at least, determine the posterior distribution of h_i. It is important to study this distribution, since it provides all the relevant information about h_i. In this section, the posterior distribution of h_i is illustrated for the one-sided cases of sampling from both normal and exponential populations.

For the normal, the joint posterior distribution of μ_i and σ_i is given by (3.3). We now make the transformation

$$h_i = g(\mu_i, \sigma_i) = \int_{-\infty}^{\frac{a_2-\mu_i}{\sigma_i}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{t^2}{2}\right\} dt = \Phi\left(\frac{a_2-\mu_i}{\sigma_i}\right)$$

(6.1)

$$v_i = \sigma_i,$$

where Φ is the cumulative distribution function of the standard normal variable.

The absolute value of the Jacobian J of the transformation (6.1) is easily seen to be given by

$$(6.2) \qquad |J| = \sqrt{2\pi}\,\sigma_i \exp\left\{\frac{(a_2-\mu_i)^2}{2\sigma_i^2}\right\}.$$

Using (6.1) and (6.2), and upon integrating over the range of v_i we find that the (marginal) posterior distribution of h_i is:

$$(6.3) \quad p(h_i|\underline{y}_i) = k_4 \exp\left\{-\frac{m_i^2}{2}\left[\frac{n_i s_i^2}{s_i^2+b_i^2} -1\right]\right\}\int_0^\infty t^{n-2} \exp\left\{-\frac{n_i(s_i^2+b_i^2)}{2}\left(t+\frac{b_i m_i}{s_i^2+b_i^2}\right)^2\right\}dt$$

where

$$k_4^{-1} = 2^{\frac{n_i-3}{2}}\, \Gamma\left(\frac{n_i-1}{2}\right) n_i^{-\frac{n_i}{2}} (s_i^2)^{-\frac{n_i-1}{2}},$$

$$m_i = \Phi^{-1}(h_i) \quad \text{and} \quad b_i = \bar{y}_i - a_2.$$

Since m_i is an inverse function of a cumulative normal, it is not possible to express the density in (6.3) in a closed form. Nevertheless, it is easy to show that this density can be put into the alternative form:

$$(6.4) \quad p(h_i|\underline{y}_i) = k_4\sqrt{2\pi}\ \exp\left\{-\frac{m_i^2}{2}\left[\frac{n_i s_i^2}{s_i^2+b_i^2} - 1\right]\right\}$$

$$[n_i(s_i^2+b_i^2)]^{-\frac{n_i-3}{2}} \sum_{p=0}^{n_i-2} \binom{n_i-2}{p}(-c_i)^{n_i-2-p}\ T_p(c_i)$$

where

$$T_p(c_i) = \frac{1}{\sqrt{2\pi}}\int_{-c_i}^{\infty} y^p\, e^{-\frac{y^2}{2}}\, dy, \text{ and } c_i = \frac{\sqrt{n_i}\ b_i m_i}{\sqrt{s_i^2+b_i^2}} .$$

The function $T_p(c_i)$ may be calculated using the following reduction formulae,*

$$p = 2r+1: \quad T_p(c) = 2^r\, r!\, \Phi'(c) \sum_{\ell=0}^{r} \frac{1}{\ell!}\left(\frac{c^2}{2}\right)^{\ell}$$

(6.5)

$$p = 2r: \quad T_p(c) = 2^{(r-\frac{1}{2})}(r-\tfrac{1}{2})!\, \Phi'(c) \sum_{\ell=1}^{r} \frac{1}{(\ell-\frac{1}{2})!}\left(\frac{c^2}{2}\right)^{\ell-\frac{1}{2}}$$

$$+ \frac{2^r(r-\frac{1}{2})!}{(\frac{1}{2}-1)!}[1-\Phi(c)] .$$

Thus, making use of an electronic computer, and consulting those standard normal tables which would permit accurate inverse interpolation, we would be able to calculate the density $p(h_i|\underline{y}_i)$. Once this is calculated, the posterior probability integrals can then be approximated using standard numerical methods, for example, Simpson's Rule, etc.

For the exponential distribution as defined by (4.1), we have found that, when y_{i1} is less than "a", the posterior distribution of μ_i and σ_i is given by expression (4.5). We now make the transformation

$$h_i = g(\mu_i, \sigma_i) = 1 - \exp\left\{-\frac{a-\mu_i}{\sigma_i}\right\}$$

(6.6)

$$v_i = \sigma_i .$$

*An equivalent form of this reduction formula is given in Fisher (1961).

The absolute value of the Jacobian of the transformation (6.5) is $|J| = \frac{v_i}{1-h_i}$. Thus, the (marginal) posterior distribution of h_i is given by

$$(6.7) \qquad p(h_i|\underline{y}_i) = \frac{n_i[(n_i-1)w_i]^{n_i-1}}{\Gamma(n_i-1)} (1-h_i)^{n_i-1} \int_{\alpha(h_i)}^{\infty} v_i^{-n_i} \exp\left\{-\frac{n_i}{v_i}[\bar{y}_i-a]\right\} dv_i$$

where

$$\alpha(h_i) = -\frac{a-y_{i1}}{\log(1-h_i)} .$$

For values of "a" different from $\bar{y}_i$ and upon making the transformation $x = n_i(\bar{y}_i - a)\, v_i^{-1}$, the integral in (6.7) can be written:

$$(6.8) \qquad [n_i(\bar{y}_i - a)]^{-(n_i-1)} \int_0^{d_i} x^{n_i-2}\, e^{-x}\, dx$$

with

$$d_i = \frac{n_i(\bar{y}_i - a)}{\alpha(h_i)} .$$

On integrating by parts, (6.8) takes the form

$$(6.9) \qquad (n_i-2)!\,[n_i(\bar{y}_i-a)]^{-(n_i-1)}\left\{1 - \exp(-d_i)\sum_{r=0}^{n_i-2} \frac{d_i^r}{r!}\right\} .$$

Using (6.9) we then have the alternative form for the posterior distribution of h_i,

$$(6.10) \qquad p(h_i|\underline{y}_i) = n_i\left[\frac{(n_i-1)w_i}{n_i(\bar{y}_i-a)}\right]^{(n_i-1)} (1-h_i)^{n_i-1}\left\{1 - \exp(-d_i)\sum_{r=0}^{n_i-2} \frac{d_i^r}{r!}\right\} .$$

From (6.10) it is easily shown that the posterior probability

$$(6.11) \qquad \Pr(h_i > h_{io}|\underline{y}_i) =$$

$$\left[\frac{(n_i-1)w_i}{n_i(\bar{y}_i-a)}\right]^{n_i-1} (1-h_{io})^{n_i} - \frac{n_i(a-y_{i1})}{(n_i-1)w_i} \exp\left\{-u_{io}\right\}\sum_{r=0}^{n_i-2}\sum_{v=0}^{r} \frac{u_{io}^v}{v!}\left[\frac{n_i(\bar{y}_i-a)}{(n_i-1)w_i}\right]^{r-n_i+}$$

where $u_{io} = \frac{(n_i-1)w_i}{\alpha(h_{io})}$. We note that expressions (6.10) and (6.11) are amenable to calculation.

As an example of the resulting posterior distribution for the exponential case discussed above, three samples of size ten each from exponential populations with $(\mu_i, \sigma_i) = (0, 1)$, $(.1, 1.1)$, and $(.2, 1.2)$ were drawn. The sample information is summarized as follows:

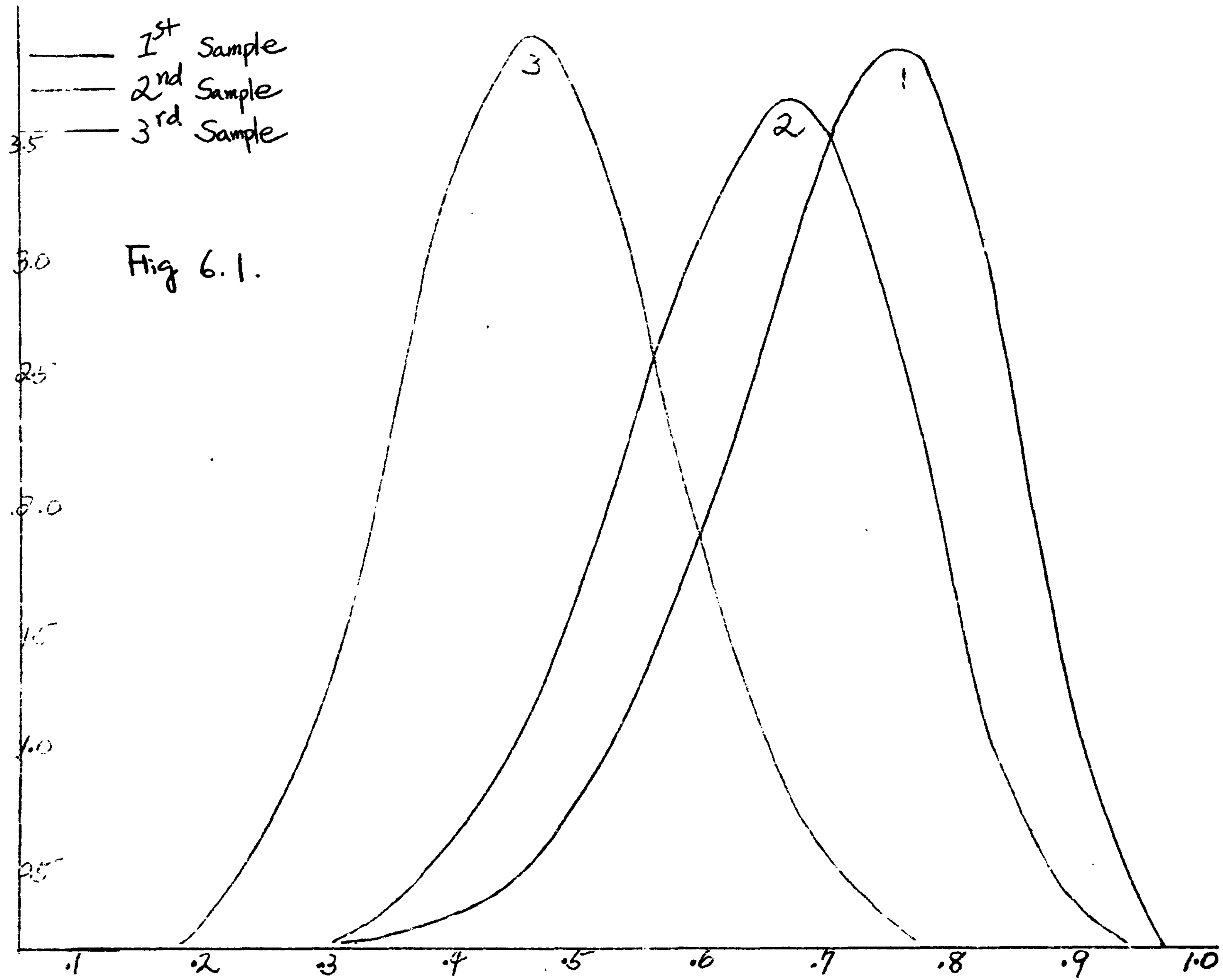
Sample No.	1	2	3
y_{i1}	.0316	.2286	.5799
w_i	.9238	.9744	1.136
$\bar{y}_i$	.8630	1.106	1.602

The value of "a" chosen is 1.20. The resulting posterior distributions are plotted in Figure 6.1.

We see from the diagrams that the posterior distribution for population 3 is nearly symmetric. However, the posterior distributions for populations 1 and 2 are skewed, the former being more so in this respect than the latter.

In fact, on examining the moments of h_i given in (5.4), it is evident that we should expect some skewness of the above type. It is for this reason that a desire for a "conservative" decision rule might make itself felt. For example, we might find that the situation depicted in Figure 6.2 might obtain.

Here we have found that the expected values of h_i $(i = 1, 2)$ are nearly the same, yet as we are interested in picking the best population, that is, the population with the largest value of h_i, we are inclined to favor population 1 since the a posteriori probability of h_1 exceeding its expected value is obviously larger than the a posteriori probability of h_2 exceeding this value. This leads us to propose the following procedure which may be favored by some experimenters. namely,



1st Sample
2nd Sample
3rd Sample
3.5
3.0
2.5
2.0
1.5
1.0
0.5
·1
·2
·3
·4
·5
·6
·7
·8
·9
1·0
3
2
1

Fig 6.1.

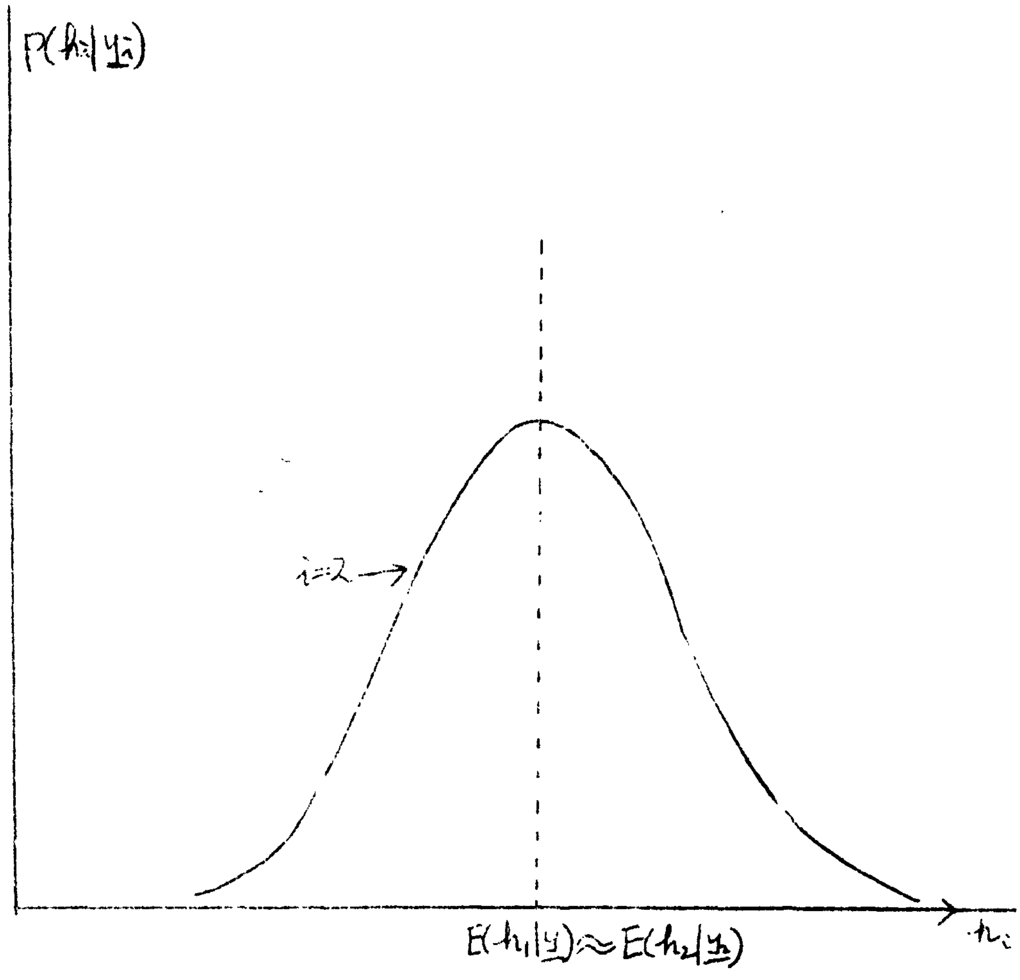


Fig. 6.2
Posterior Distributions of h_i for Two Populations having Same Expected Value.

(i) As in sections 3 and 4, find $\max_i E(h_i|\underline{y}_i)$. Let this value be denoted by γ_o.

(ii) Determine the a posteriori probability that h_i exceeds the quantity γ_o, say P_i, for each population. Make the decision that the population having as its value of P, the value $\max_i P_i$, is the best population.

Other procedures of this type suggest itself at this point. For example, suppose the statistician is concerned with a mass production process in which items are defined to be non-defective if, re a certain specification, they measure less than the quantity "a". Then obviously, among k processes producing this item, the statistician wishes to choose the process yielding the largest proportion of items which measure less than "a". Further, suppose that the processes are costly and that the break even point is accomplished if and only if the proportion of non-defectives is greater than a known percentage, say 100 γ%. Then one would expect the statistician to choose as the best population the one having largest value for the posteriori probability $P_i = \Pr[h_i > \gamma|\underline{y}_i]$.* In fact, statisticians may be faced with situations which would call for rules of this type but using different choices of γ. The value of γ chosen would depend entirely on the particular problem involved.

We remark that knowledge of the posterior distribution has enabled us to arrive at the decision rules of the types described in this section. It seems to us entirely natural to base our decision upon the posterior distribution, in preference to the rules exemplified by sections 3 and 4, which are based on maximizing the expected value

---

*We are indebted to Professor David Finney (1962) for this remark.

$E(h_i|\underline{y}_i)$. For, after all, the posterior distribution gives us all the information about h_i including its expected value. Of course, because of the consistency property of the procedures of sections 3 and 4 (as shown in section 5), we recommend that the procedures of this section be used only for small samples.

VII. Some Other Best Population Problems

The problems discussed in sections 3 and 4 involved "utility" functions $h_i = g(\theta_i)$ which are rather complicated functions of the population parameters. The analysis used in these two sections was that dictated by the general framework described in section 2. Using the same approach, we turn now to some other best population problems where the definitions of "best" used, involve h_i which are simple functions of either a location or a scale parameter of the population distributions. The specific h_i and their expectations are summarized in Table 7.1.

Table 7.1

Utility	Expected Utility When Sampling From		
$h_i=g(\theta_i)$	$\frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{1}{2\sigma_i^2}(y_i-\mu_i)^2\right\}$	$\frac{1}{\sigma_i} \exp\left\{-\frac{1}{\sigma_i}(y_i-\mu_i)\right\}$	$\mu_i^{y_i}(1-\mu_i)^{1-y_i}$
μ_i	$\bar{y}_i$	$y_{i1} - \frac{(n_i - 1)w_i}{n_i(n_i - 2)}$	$\frac{n_i\bar{y}_i + \ell_i}{n_i + \ell_i + m_i}$
$\frac{1}{\sigma_i}$	$\frac{\sqrt{2}\,\Gamma(\frac{n_i}{2})}{\sqrt{n_i}\,\Gamma(\frac{n_i-1}{2})\, s_i}$	$\frac{1}{w_i}$	-----
Prior Distributions Of Parameters	$p(\mu_i) \propto k$ $p(\sigma_i) \propto \frac{1}{\sigma_i}$	$p(\mu_i) \propto k$ $p(\sigma_i) \propto \frac{1}{\sigma_i}$	$p(\mu_i)= \frac{\Gamma(\ell_i+m_i)}{\Gamma(\ell_i)\Gamma(m_i)}$ $\mu_i^{\ell_i-1}(1-\mu_i)^{m_i-1}$

These types of best population problems summarized above have been considered from the non-Bayesian point of view -- see Gupta and Sobel (1960, 1962), Bechhofer (1954), and Bechhofer and Sobel (1954) -- and from the Bayesian point of view by Raiffa and Schlaifer (1961). It is easy to show that the results listed in Table 7.1 are consistent. Also, these problems can of course be treated using the analysis discussed in section 6.

References

Bechhofer, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Statist. 25, 16-29.

Bechhofer, R. E. and Sobel, M. (1954). A single-sample multiple decision procedure for ranking variances of normal population. Ann. Math. Statist. 25, 273-89.

Box, G. E. P. and Tiao, G. C. (1962). A further look at robustness via Bayes' theorem. Biometrika, 49, 419-32.

Dunnett, C. W. and Sobel, M. (1954) A bivariate generalization of Student's t-distribution, with tables for certain special cases. Biometrika, 41, 153-69.

Finney, D. J. (1962). Personal communication.

Fisher, R. A. (1961). The weighted mean of two normal samples with unknown variance ratio. Sankhya, 23, 103-14.

Gupta, S. S. and Sobel, M. (1960). Selecting a subset containing the best of several binomial populations. Chapter XX, pp. 224-48 in Contributions to Probability and Statistics. Stanford University Press.

Gupta, S. S. and Sobel, M. (1962). On selecting a subset containing the population with the smallest variance. Biometrika, 49, 495-507.

Guttman, Irwin (1961). Best populations and tolerance regions. Ann. of the Inst. of Statist. Math. (Tokyo), 13, 9-26.

Jeffreys, H. (1961). Theory of Probability (3rd edition). Oxford: Clarendon Press.

Luce, R. D. and Raiffa, H. (1957). Games and Decisions. New York: John Wiley and Sons.

Raiffa, H. and Schlaifer, R. (1961). Applied Statistical Decision Theory, Harvard University.

Savage, L. J. (1961) The Subjective Basis of Statistical Practice. Manuscript, University of Michigan.